AD-A086 162 STANFORD UNIV CA SYSTEMS OPTIMIZATION LAB F/G 12/1
A PROJECTED LAGRANGIAN ALGORITHM FOR NONLINEAR 'L (SUB 1)' OPTI--ETC(U)
FEB 80 W MURRAY, M L OVERTON DAAG29-79-C-0110
UNCLASSIFIED SOL-80-4 ARO-16470.10-M NL

1 OF 1
AD A086 162

END
DATE FILMED
8-80
DTIC

LEVEL II

ARO 16470.10-M

(12)

Systems
Optimization
Laboratory

ADA086162

(6) A PROJECTED LAGRANGIAN ALGORITHM FOR NONLINEAR ℓ_1 OPTIMIZATION

by

Walter Murray and Michael L. Overton*

(9) TECHNICAL REPORT SOL 80-4
Feb 80
(11)

DDC FILE COPY

Department of Operations Research
Stanford University
Stanford, CA 94305

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
Stanford University
Stanford, California
94305

(15) DAA[illegible]
DE-AC03-[illegible]

(6) A PROJECTED LAGRANGIAN ALGORITHM
FOR NONLINEAR ℓ_1 OPTIMIZATION

by

(1) Walter Murray and Michael L. Overton*

(9) TECHNICAL REPORT SOL 80-4
Feb[illegible]80

(11)

(12) 49

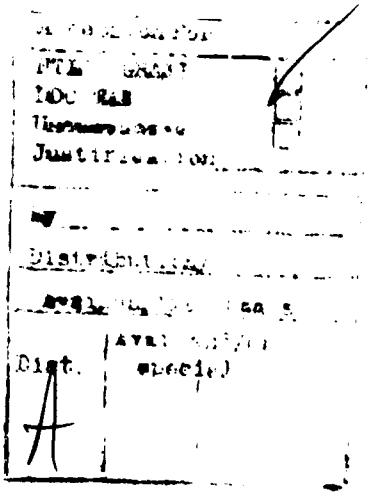
(14) SOL-80-4

Research and reproduction of this report were partially supported by the Department of Energy Contract DE-AC03-76-SF00326, PA Number DE-AT-03-76ER72018; the National Science Foundation Grant MCS-7926009; and the Army Research Office Contract DAAG29-79-C-0110.

*Research and reproduction of this report were partially supported by the Department of Energy Contract DE-AS03-76SF00326, PA Number 30; and the National Science Foundation Grant MCS78-11985, Computer Science Department, Stanford University.

ABSTRACT

The nonlinear ℓ_1 problem is an unconstrained optimization problem whose objective function is not differentiable everywhere, and hence cannot be solved efficiently using standard techniques for unconstrained optimization. The problem can be transformed into a nonlinearly constrained optimization problem, but it involves many extra variables. We show how to construct a method based on projected Lagrangian methods for constrained optimization which requires successively solving quadratic programs in the same number of variables as that of the original problem. Special Lagrange multiplier estimates are used to form an approximation to the Hessian of the Lagrangian function, which appears in the quadratic program. A special line search algorithm is used to obtain a reduction in the ℓ_1 objective function at each iteration. Under mild conditions the method is locally quadratically convergent if analytical Hessians are used.



A PROJECTED LAGRANGIAN ALGORITHM FOR NONLINEAR ℓ_1 OPTIMIZATION

Walter Murray and Michael L. Overton

1. Introduction

The problem we wish to solve is

$$\ell_1 P : \min_x \{F_1(x) \mid x \in R^n\}$$

$$\text{where} \qquad F_1(x) = \sum_{i=1}^{m} |f_i(x)|$$

and the functions $f_i : R^n \to R^1$ are twice continuously differentiable. The function $F_1(x)$ is called the ℓ_1 function and $\ell_1 P$ is referred to as the ℓ_1 problem. The ℓ_1 problem is an unconstrained optimization problem in which the objective function has discontinuous derivatives and hence it is inappropriate to use a standard unconstrained minimization method to solve it. The problem is equivalent to the following nonlinearly constrained problem in which both the objective and constraint functions are twice continuously differentiable:

$$\text{ELP} : \min\{ \sum_{i=1}^{m} u_i \mid x \in R^n,\ u \in R^m\}$$

$$\text{subject to} \quad c_i^{(\sigma)}(x,u) \geq 0,\ i = 1,2,\ldots,m;\ \sigma = -1,1,$$

$$\text{where} \qquad c_i^{(\sigma)}(x,u) = u_i - \sigma f_i(x) \ .$$

We could solve ELP using a method for the general nonlinear programming problem, but this is very unattractive since m, the number of extra variables, may be large. A method can be derived which exploits the special structure of problem ELP essentially reducing it back to a problem with n variables. One special feature of ELP is that the ℓ_1 function F_1 is a natural merit function which can be used to measure progress towards the solution of ELP. Such a merit function is not generally available for the nonlinear programming problem without the introduction of a parameter such as penalty parameter.

The method we adopt to solve $\ell_1 P$ consists of two parts at each iteration: (1) obtain a direction of search by solving and perhaps modifying a quadratic program based on a projected Lagrangian algorithm for ELP, and (2) take a step along the search direction which reduces the ℓ_1 function. The general approach is similar to that described for the minimax problem by Murray and Overton [15]. The structure of the quadratic program to be solved is however considerably different from the minimax case and this is described in full in subsequent sections. We use a special line search algorithm which is closely related to the one used in the minimax case. This is discussed in Section 7; the details may be found in [14].

A number of other algorithms have been proposed for solving the nonlinear ℓ_1 problem. These will be discussed further in Section 9, after our algorithm has been described in full. At the time of this writing, no other algorithms related to projected Lagrangian methods

using second order information have, to our knowledge, been published. However, Bartels and Conn are currently doing some related work.

We note that no convexity assumptions are made about the functions $f_i(x)$. We concern ourselves only with local minima.

1.1. Notation

All vectors are column vectors, but for convenience we will write (x,u) for $\begin{pmatrix} x \\ u \end{pmatrix}$. Define $(\overset{*}{x},\overset{*}{u})$ to be a solution of ELP. It follows that $\overset{*}{x}$ is a solution to $\ell_1 P$ and

$$F_1(\overset{*}{x}) = \sum_{i=1}^{m} \overset{*}{u}_i .$$

Let $(x^{(k)},u^{(k)})$ denote the k-th approximation to $(\overset{*}{x},\overset{*}{u})$.

At each iteration of the algorithm $(x^{(k+1)},u^{(k+1)})$ is obtained by setting

$$x^{(k+1)} = x^{(k)} + \alpha p \quad \text{and} \quad u^{(k+1)} = |f(x^{(k+1)})|$$

where p is the direction of search in R^n and α, a positive scalar, is the steplength, and the absolute value of a vector denotes the vector of the absolute values of the components. Note that this choice of $(x^{(k+1)}, u^{(k+1)})$ immediately guarantees that all the points $\{(x^{(k)},u^{(k)})\}$ are <u>feasible</u> for ELP, i.e. $c_i^{(\sigma)}(x^{(k)}) \geq 0$, $i = 1,\ldots,m$, $\sigma = -1,+1$. It also follows that for each i, at least one of the pair of constraints $(c_i^{(-1)}, c_i^{(+1)})$ must have the value zero. We will be interested in the case where the other constraint in the pair is also zero at the

solution, i.e. the corresponding function f_i is zero. Therefore at any point x we define the <u>active set of functions</u> as those which we think will have the value zero at the solution $\overset{*}{x}$, based on the information at x. This set will usually include all functions with the value zero at the point x and may also include some with nonzero values. The exact procedure for selecting the active set at each iteration will be discussed in Section 8 and procedures for modifying this choice will be discussed in Section 5.

We define t $(= t(x))$ to be the number of active functions at x and write the vector of active functions as $\hat{f}(x) \in R^t$. Define $\hat{\Sigma}(x)$ to be the diagonal square matrix of order t whose $i^{\underline{th}}$ diagonal component is 1 if $\hat{f}_i(x) \geq 0$ and -1 otherwise. Define $\hat{V}(x)$ to be the $n \times t$ matrix whose columns $\{\hat{v}_i(x)\}$ are the gradients of the active functions. Similarly we define $\bar{f}(x) \in R^{m-t}$ to be the vector of inactive functions at x and define $\bar{\Sigma}(x)$ and $\bar{V}(x)$ to be respectively the $(m-t) \times (m-t)$ diagonal matrix of the signs corresponding to $\bar{f}(x)$ and the $n \times (m-t)$ matrix of gradients of the inactive functions. We also define $\bar{u}$ and $\hat{u}$ to be the subvectors of u corresponding to $\bar{f}$ and $\hat{f}$.

We define the <u>active constraints</u> at x to be <u>both</u> constraints of each pair corresponding to the active functions plus the one constraint with zero value of each pair corresponding to the inactive functions. We can order the active constraints so that the vector of active constraint values is given by

$$\hat{c}(x) = \begin{bmatrix} \bar{u} - \bar{\Sigma}(x)\ \bar{f}(x) \\ \hat{u} - \hat{f}(x) \\ \hat{u} + \hat{f}(x) \end{bmatrix}$$

with $\bar{u} = \bar{\Sigma}(x)\ \bar{f}(x)$, $\hat{u} = \hat{\Sigma}(x)\ \hat{f}(x)$, by definition of u. Define $\hat{A}(x)$ to be the $(m+n) \times (m+t)$ matrix whose columns $\{\hat{a}_j\}$ are the active constraint gradients. We can order the variables $\{u_i\}$ so that

$$\hat{A}(x) = \begin{bmatrix} -\bar{V}(x)\ \bar{\Sigma}(x) & -\hat{V}(x) & \hat{V}(x) \\ I_{m-t} & 0 & 0 \\ 0 & I_t & I_t \end{bmatrix}$$

Here I_s is the identity matrix of order s.

We define $Y(x)$ *and* $Z(x)$ to be orthogonal matrices respectively spanning the range and null spaces of $\hat{V}(x)$. Provided $\hat{V}(x)$ has full rank we have that $Y(x)$ has dimension $n \times t$, $Z(x)$ has dimension $n \times (m-t)$, and

$$Y(x)^T\, Y(x) = I_t, \qquad Z(x)^T\, Z(x) = I_{n-t},$$
$$Y(x)^T\, Z(x) = \hat{V}(x)^T\, Z(x) = 0 .$$

Let g be the gradient of the objective function of ELP, i.e. the $(n+m)$-vector:

$$g = \begin{pmatrix} 0 \\ \bar{e} \\ \hat{e} \end{pmatrix}$$

where $\bar{e} \in R^{m-t}$ and $\hat{e} \in R^t$ are vectors of all ones.

The Lagrangian function associated with ELP is

$$L(x,u,\lambda) = \bar{e}^T\bar{u} + \hat{e}^T\hat{u} - \lambda^T\hat{c}(x)$$

where $\lambda \in R^{m+t}$ is a vector of Lagrange multipliers. The gradient of $L(x,u,\lambda)$ with respect to x is $g - \hat{A}\lambda$. Define W_E to be the Hessian of the Lagrangian function with respect to (x,u). Then

$$W_E = \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix}$$

where W is the Hessian of $L(x,u,\lambda)$ with respect to x only, i.e.

$$W(x,\lambda) = \sum_{i=1}^{m+t} \lambda_i \nabla_x^2 \hat{c}_i(x) \ .$$

Define Z_E to be an orthogonal matrix spanning the null space of $\hat{A}$, i.e., $\hat{A}^T Z_E = 0$. It follows from the definition of $\hat{A}$ that the first n rows of Z_E can be taken to be Z, the matrix which is orthogonal to $\hat{V}$. Thus $Z_E^T W_E Z_E$, the Hessian of $L(x,u,\lambda)$ projected into the null space of $\hat{A}$, can also be written as $Z^T W Z$.

We will use p_E to denote a vector in R^{n+m} whose first n components are the direction of search vector p. We write

$$p_E = \begin{pmatrix} p \\ \bar{p} \\ \hat{p} \end{pmatrix}$$

where $\bar{p}$ and $\hat{p}$ correspond to $\bar{u}$ and $\hat{u}$.

Often we will omit the arguments from the various vectors and matrices $\hat{f}$, $\bar{V}$, $\hat{A}$ etc. when it is clear that they are evaluated at $x^{(k)}$. We use the notation $\overset{*}{A}$, $\overset{*}{V}$, $\overset{*}{Z}$ to denote $\hat{A}$, $\hat{V}$, Z evaluated at $(\overset{*}{x},\overset{*}{u})$ with the active set of functions correctly chosen, i.e., consisting of all those functions with the value zero at $\overset{*}{x}$.

1.2. Necessary and Sufficient Conditions

The necessary and sufficient conditions for $(\overset{*}{x},\overset{*}{u})$ to be a local minimum of problem ELP and therefore $\overset{*}{x}$ to be a local minimum of problem $\ell_1 P$ are simplifications of the general necessary and sufficient conditions for the nonlinear programming problem. A similar argument to that given in [15] shows that the first-order constraint qualification always holds for ELP. The conditions therefore reduce to the following:

First-order necessary condition

If $(\overset{*}{x},\overset{*}{u})$ is a local minimum of ELP then there exists a vector of Lagrange multipliers $\overset{*}{\lambda} \in R^{m+t}$ such that

$$g - \overset{*}{A}\overset{*}{\lambda} = 0 \qquad \text{and} \qquad \overset{*}{\lambda} \geq 0 \ . \tag{1.2}$$

Second-order necessary condition

If $(\overset{*}{x},\overset{*}{u})$ is a local minimum of ELP and the second-order constraint qualification holds, then $\overset{*}{Z}^T W(\overset{*}{x},\overset{*}{\lambda})\overset{*}{Z}$, the projected Hessian of the Lagrangian function, is positive semi-definite.

Sufficient condition

If the first-order necessary condition holds at $(\overset{*}{x},\overset{*}{u})$, the Lagrange multipliers are all strictly positive, i.e. $\overset{*}{\lambda} > 0$, and $\overset{*}{Z}^T W(\overset{*}{x},\overset{*}{\lambda})\overset{*}{Z}$ is positive definite, then $(\overset{*}{x},\overset{*}{u})$ is a strong local minimum of ELP. Thus in terms of $\ell_1 P$, $F_1(\overset{*}{x}) < F_1(x)$ for all x such that $|x-\overset{*}{x}| < \delta$, for some $\delta > 0$.

In the case where all the $\{f_i\}$ are linear it is well-known that a solution must exist with n active functions at $\overset{*}{x}$. Then normally the matrix $\overset{*}{Z}$ is null which implies the second-order conditions are also null. The nonlinear problem, however, can have a unique solution with anything from zero to n functions active at $\overset{*}{x}$.

2. Use of the Equivalent Problem ELP

At every iteration we wish the search direction p to be a descent direction for F_1, i.e.

$$F_1'(x^{(k)},p) < 0 ,$$

where $F_1'(x^{(k)},p)$ is the directional derivative

$$\lim_{h \to 0+} \frac{1}{h} (F_1(x^{(k)} + hp) - F_1(x^{(k)})) .$$

It is easy to see that $F_1'(x^{(k)},p)$ is also given by

$$\sum_{i|f_i\neq 0} \frac{f_i}{|f_i|} v_i^T p + \sum_{i|f_i=0} |v_i^T p| \qquad (2.1)$$

We have the following:

THEOREM 1. If p_E is a first-order feasible descent direction for ELP, i.e.

$$g^T p_E < 0 \qquad (2.2)$$

and

$$Vc_i^{(\sigma)^T} p_E \geq 0 \quad \text{for all} \quad i, \sigma \quad \text{such that} \quad c_i^{(\sigma)} = 0 , \qquad (2.3)$$

then p is a descent direction for F_1 and hence a sufficiently small step along it must result in a reduction in F_1.

Proof. Suppose for the moment that the active set consists of those and only those functions which are zero at $x^{(k)}$, so that we can use the notation developed for this. We then have $\hat{A}^T p_E \geq 0$, and hence

$$-\bar{\Sigma}\bar{V}^T p + \bar{p} \geq 0$$

$$-\hat{V}^T p + \hat{p} \geq 0$$

$$\hat{V}^T p + \hat{p} \geq 0 .$$

It follows from (2.1) and (2.2) that

$$F_1'(x^{(k)}, p) \leq \bar{e}^T\bar{p} + \hat{e}^T\hat{p} < 0 . \qquad \square$$

It is possible for p_E to be a first-order feasible direction without being a feasible direction for ELP. This causes no difficulty since $u^{(k+1)}$ is set to $|f(x^{(k+1)})|$ and hence it is always possible to obtain a lower feasible point for ELP if (2.2) and (2.3) hold, by reducing F_1 along p.

A second desirable property for p arises from considering the active set of functions, i.e. those we expect to have the value zero at $\overset{*}{x}$. We wish to choose p so that the first-order change in these functions predicts that they will all have the value zero at $x^{(k)} + p$. An equivalent condition is:

$$\hat{V}^T p = -\hat{f} \ . \tag{2.4}$$

This condition is implied by the following condition on the $(n+m)$-vector p_E:

$$\hat{A}^T p_E = -\hat{c} \ . \tag{2.5}$$

Strictly speaking, (2.5) is a stronger condition than the pair of conditions (2.4) and

$$\hat{A}^T p_E \geq -\hat{c} \ . \tag{2.6}$$

However, since the only difference is that the variables $\{u_i\}$ are required to be on their bounds and it will become evident later that this does not affect the choice of search direction, for simplicity we will require that p_E satisfy (2.5).

Thus we see that one view of ELP is as a device to obtain a search direction p along which F_1 can be reduced in the line search.

We emphasize again that we wish (2.2) and (2.3) to hold so that p is a descent direction for F_1, and that the active set nature of the algorithm indicates that (2.4) and hence (2.5) should also hold.

3. Derivation and Solution of QP Subproblem

The solution of ELP is at a minimum of the Lagrangian function in the null space of the active constraint Jacobian. The usual method for solving a general linearly constrained problem is to approximate the objective function by a quadratic function and then determine the search direction by solving some appropriate quadratic program (QP). Consider therefore the quadratic program:

$$\text{QP1:} \qquad \min_{p_E} \frac{1}{2} p_E^T W_E(x^{(k)}, \lambda^{(k)}) p_E + g^T p_E$$

$$\text{subject to } \hat{A}^T p_E = -\hat{c} \ ,$$

$$\text{where } \lambda^{(k)} \text{ is an estimate of } \overset{*}{\lambda} \ .$$

The constraints of QP1 are equivalent to (rearranging equations):

$$\begin{aligned} -\bar{\Sigma}\bar{V}^T p + \bar{p} &= 0 \\ -\hat{\Sigma}\hat{V}^T p + \hat{p} &= 0 \\ \hat{\Sigma}\hat{V}^T p + \hat{p} &= -2\hat{\Sigma}\hat{f} \end{aligned} \qquad (3.1)$$

The last two equations imply that

$$\hat{p} = \hat{\Sigma}\hat{V}^T p = -\hat{\Sigma}\hat{f} \tag{3.2}$$

Since $g^T p_E = \bar{e}^T \bar{p} + \hat{e}^T \hat{p}$ it follows that p can be obtained by solving the following QP in only n variables:

$$\text{QP2:} \qquad \min_p \; p^T W(x^{(k)}, \lambda^{(k)}) p + \bar{e}^T \bar{\Sigma} \bar{V}^T p$$

$$\text{subject to } \hat{V}^T p = -\hat{f} \; .$$

In order to solve QP2 we introduce the matrices Y and Z defined in Section 1.1. These provide bases for the range and null spaces of $\hat{V}$. The matrices Y and Z may be determined from the QR factorization of $\hat{V}$:

$$\hat{V} = Q \begin{bmatrix} R \\ 0 \end{bmatrix} = [Y \quad Z] \begin{bmatrix} R \\ 0 \end{bmatrix}$$

where R is an upper triangular matrix of order t. If $\hat{V}$ has full rank and $Z^T W Z$ is positive definite, then the unique solution of QP2 may be expressed as the sum of two orthogonal components

$$p = Y p_Y + Z p_Z \tag{3.3}$$

where $p_Y \in R^t$ and $p_Z \in R^{n-t}$. We have

$$\hat{V}^T p = R^T p_Y = -\hat{f} \tag{3.4}$$

and p_Y is determined entirely by the constraints of QP1. The vector p_Z is given by the solution of

$$(Z^T W Z) p_Z = -Z^T(\bar{\bar{V}}\bar{\bar{\Sigma}}\bar{\bar{e}} + WYp_Y) \tag{3.5}$$

We shall also wish to refer to the related QP with homogeneous constraints:

$$\text{QP3:} \qquad \min_p \ \frac{1}{2} p^T W p + \bar{\bar{e}}^T \bar{\bar{\Sigma}} \bar{\bar{V}}^T p$$

$$\text{subject to} \ \ \hat{V}^T p = 0 \ .$$

The solution to this is given by $p = Zq_z$, where

$$(Z^T W Z) q_z = -Z^T \bar{\bar{V}}\bar{\bar{\Sigma}}\bar{\bar{e}} \ . \tag{3.6}$$

At every iteration of our algorithm an attempt is made to set the search direction p to the solution of QP2, but for various reasons this may be inadequate. In subsequent sections we explain what action is taken in these circumstances.

4. Lagrange Multiplier Estimates

Let us first suppose that $x^{(k)}$ is a minimum on the manifold defined by the current active set. Then for some λ we have

$$\hat{A}\lambda = g.$$

Writing $\lambda = (\bar{\lambda}, \hat{\lambda}^+, \hat{\lambda}^-)$ this is equivalent to

$$-\bar{V}\bar{\Sigma}\bar{\lambda} - \hat{V}\hat{\lambda}^+ + \hat{V}\hat{\lambda}^- = 0 \tag{4.1}$$

$$\bar{\lambda} = \bar{e} \tag{4.2}$$

$$\hat{\lambda}^+ + \hat{\lambda}^- = \hat{e} \ . \tag{4.3}$$

Let us therefore define

$$\pi = \hat{\lambda}^+ - \hat{\lambda}^- \ . \tag{4.4}$$

Equations (4.1) and (4.2) then reduce to

$$\hat{V}\pi = -\bar{V}\bar{\Sigma}\bar{e} \ . \tag{4.5}$$

It follows from (4.3) and (4.4) that the first order necessary condition $\overset{*}{\lambda} \geq 0$ is equivalent to

$$|\overset{*}{\pi}| \leq 1$$

where $\overset{*}{\pi}$ is π at $\overset{*}{x}$. The vector π here plays the same role as the vector w in the linear case described by Bartels, Conn and Sinclair [5].

The question we face in this section is how to define the vector of Lagrange multiplier estimates at any point $x^{(k)}$, dropping the assumption that it is the minimum on the manifold. Multiplier estimates are needed to define the matrix W and to determine whether constraints should be deleted from the active set. It is better to use new information obtained at the current point $x^{(k)}$ rather than use the multipliers of the QP solved at the previous iteration. Such an estimate should in some sense approximately satisfy the overdetermined system based on the first order necessary conditions:

$$\hat{A}\lambda \approx g \ . \tag{4.6}$$

Clearly it makes no sense to delete an active constraint corresponding to an inactive function, since the corresponding variable $\bar{u}_i$ will be reduced at the end of the iteration to make the constraint active again. Similarly, only one active constraint of the pair corresponding to an active function should be considered for deletion. These facts combined with the fact that the search direction is being determined for a QP involving only n variables indicate that a special estimate taking into account the structure of ELP should be used, as opposed to the least squares solution of (4.6).

λ_C: The special estimate for ELP

The special estimate is required to satisfy exactly those equations in (4.6) which are exactly the same as the equations holding at the minimum on the manifold defined by the active set. Thus we define λ_C to be the least squares solution to the approximate equation (4.1), subject to the constraint that (4.2) and (4.3) hold exactly. Equivalently we can define π_C as the solution of the least squares problem

$$\min \|\hat{V}\pi + \bar{V}\bar{\Sigma}\bar{e}\|_2^2 . \tag{4.7}$$

It is then not necessary to explicitly form λ_C, since checking whether a component of π is greater than one in modulus is equivalent to checking whether a component of $\hat{\lambda}^+$ or $\hat{\lambda}^-$ is negative. Furthermore, using λ_C to define W results in

$$W = \sum_{i=1}^{m-t} (\bar{\lambda}_C)_i \bar{\sigma}_i \nabla^2 \bar{f}_i + \sum_{i=1}^{t} (\hat{\lambda}_C^+)_i \nabla^2 \hat{f}_i + \sum_{i=1}^{t} -(\hat{\lambda}_C^-)_i \nabla^2 \hat{f}_i = \sum_{i=1}^{m-t} \bar{\sigma}_i \nabla^2 \bar{f}_i + \sum_{i=1}^{t} (\pi_C)_i \nabla^2 \hat{f}_i,$$

where $\bar{\sigma}_i$ is the i^{th} diagonal element of $\bar{\Sigma}$. Thus π_C may also be used to define W directly.

The vector π_C can be computed directly from the QR factorization of $\hat{V}$ which we introduced in the last section to solve QP2. The estimate is a first-order multiplier estimate in the sense defined in [15].

Because at every iteration computing the search direction involves only the first n variables of ELP, the multiplier estimate which is relevant to predicting whether the steepest descent step in a subspace of R^n will be first-order feasible is λ_C, not the least squares solution to (4.6). Let us suppose that $(\pi_C)_j > 1$ and that the constraint $\hat{u}_j \geq -\hat{f}_j(x)$ is to be deleted from the active constraint set (i.e. $\hat{f}_j$ is to be deleted from the active function set). Define $\tilde{V}$ as $\hat{V}$ with $\hat{v}_j$ deleted, and $\tilde{Z}$ by

$$\tilde{V}^T\tilde{Z} = 0, \quad \tilde{Z}^T\tilde{Z} = I_{n-t+1}, \quad \tilde{Z} = [Z \quad z] . \tag{4.8}$$

Now consider the gradient of the linear term in QP2, i.e., $\bar{V}\bar{\Sigma}\bar{e}$. Since $\hat{v}_j$ is being deleted from the active set it should be included in the gradient of an objective function to be minimized in the null space of $\tilde{V}$. Therefore define the steepest descent step in the new null space to be

$$\underset{Z}{\tilde{z}s} = -\tilde{Z}\tilde{Z}^T(\bar{V}\bar{\Sigma}\bar{e} + \hat{v}_j) .$$

(Here $\hat{v}_j$ has a positive sign since $(\pi_C)_j > 1$.) We then have the following result relating the estimate π_C to the first-order feasibility of $\tilde{Z}s_{\tilde{Z}}$.

THEOREM 2. Assume $\hat{V}$ has full rank. If $(\pi_C)_j > 1$, then

$$\hat{v}_j^T \tilde{Z} s_{\tilde{Z}} > 0 .$$

Proof. We have the least squares characterization:

$$\hat{V}\pi_C = -(I - ZZ^T)\bar{\bar{V}}\bar{\bar{\Sigma}}\bar{\bar{e}} .$$

Thus

$$(z^T\hat{v}_j)(\pi_C)_j = -z^T\bar{\bar{V}}\bar{\bar{\Sigma}}\bar{\bar{e}} .$$

By definition of $\tilde{Z}s_{\tilde{Z}}$ we have:

$$\begin{aligned}
\hat{v}_j^T \tilde{Z} s_{\tilde{Z}} &= \hat{v}_j^T [Z \quad z]\, s_{\tilde{Z}} \\
&= -\,[0 \quad \hat{v}_j^T z] \begin{bmatrix} Z^T \\ z^T \end{bmatrix} (\bar{\bar{V}}\bar{\bar{\Sigma}}\bar{\bar{e}} + \hat{v}_j) \\
&= -(\hat{v}_j^T z)\,(z^T\bar{\bar{V}}\bar{\bar{\Sigma}}\bar{\bar{e}} + z^T\hat{v}_j) \\
&= -(\hat{v}_j^T z)^2\,(1 - (\pi_C)_j) > 0 .
\end{aligned}$$

(The fact that $\hat{V}$ has full rank implies that $\hat{v}_j^T \tilde{Z} \neq 0$ and hence $\hat{v}_j^T z \neq 0$.) □

It follows from Theorem 2 that setting $p = \tilde{Z}s_{\tilde{Z}}$ defines a vector p_E which is first-order feasible w.r.t. the deleted constraint

$\hat{u}_j \geq -\hat{f}_j(x)$, since p_E is first-order feasible w.r.t. the retained active constraint $\hat{u}_j \geq \hat{f}_j(x)$. Note that $(\pi_C)_j$ being > 1 is equivalent to $(\hat{\lambda}_C^-)_j$, the multiplier corresponding to the deleted constraint, being negative. Clearly if $(\pi_C)_j < -1$ and hence $(\hat{\lambda}_C^+)_j < 0$ then the deleted constraint would be $\hat{u}_j \geq \hat{f}_j(x)$, the steepest descent step would be $-\tilde{Z}\tilde{Z}^T(\bar{V}\bar{\Sigma}\bar{e} - \hat{v}_j)$, and this would have a negative inner product with $\hat{v}_j$.

λ_L: The least squares estimate in the larger space.

Define λ_L to be the least squares solution to (4.6). It is worth emphasizing that although the signs of the components of λ_L determine the feasibility of steepest descent steps for ELP in the null space of $\hat{A}$ with a column deleted (see [10] and [15]), the estimate λ_L is not relevant to the algorithm we describe for solving the ℓ_1 problem. This is because (unlike in the minimax case of [15]), it is the range and null spaces of $\hat{V}$, not $\hat{A}$, which determine the search direction at each iteration. Unlike the minimax case, λ_L and λ_C are not scalar multipliers of each other. It will often be the case that a component of λ_C is negative while the corresponding component of λ_L is positive, indicating that if the corresponding constraint is deleted, the steepest descent step (with respect to ELP) in the new null space in the $(n+m)$-dimensional space will not be feasible with respect to the deleted constraint, while the steepest descent step (with respect to QP2) in the new null space in the n-dimensional space will be first-order feasible. The converse is much less likely to happen,

i.e., where a component of λ_C is positive while the corresponding component of λ_L is negative, but it is possible to construct such an example.

Quite apart from its deficiencies, the estimate λ_L is more expensive to compute than λ_C, since it involves the factorization of a larger matrix. If m is large compared to n, then the additional effort may be prohibitively high. There is a slight simplification however: since the last t rows of $\hat{A}$ have full rank and are orthogonal to the others, it follows that the last t equations in (4.6) must hold exactly and hence a least squares problem of slightly reduced dimension can be solved.

Clearly it is undesirable to compute λ_L and we will not discuss this estimate any further.

μ_W: A Second-Order Estimate

A second-order multiplier estimate can be defined as the solution to the consistent set of overdetermined equations

$$\hat{A}\mu_W = g + W_E p_E \ .$$

The fact that the system is consistent implies that μ_W can be obtained from solving

$$\hat{V}\pi = -\bar{V}\bar{\Sigma}\bar{e} - Wp$$

and there is no concern about solving the larger system. A negative component of μ_W does not guarantee that either a steepest descent or Newton step will be first-order feasible with respect to the deleted constraint.

Using the estimate π_C to define W

Both the estimates λ_C and μ_W will be used to decide when to delete functions from the active set. Since a function corresponding to $|(\pi_C)_j| > 1$ will not necessarily be deleted from the active set, we note here that we use π_C to define W as follows:

$$W = \sum_{i=1}^{t} \pi_i' \nabla^2 \hat{f}_i + \sum_{i=1}^{m-t} \bar{\sigma}_i \nabla^2 \bar{f}_i$$

where

$$\pi_i' = \begin{cases} -1 & \text{if } (\pi_C)_i < -1 \\ 1 & \text{if } (\pi_C)_i > 1 \\ (\pi_C)_i & \text{otherwise.} \end{cases}$$

5. Properties of Solution of QP Subproblem

In this section we examine the properties of the solution to QP2. Initially, we assume that all functions with zero value are included in the active set, and that $\hat{V}$ has full rank and $Z^T W Z$ is positive definite so that the solution p is given by (3.3), (3.4) and (3.5) is unique. The corresponding solution to QP1 is p_E, where $\bar{p}$ and $\hat{p}$ are given by (3.1) and (3.2). We would like p_E to

satisfy (2.2), (2.3) and (2.5). Clearly the constraints of QP1 ensure that (2.3) and (2.5) hold. Thus the only question is whether p_E is a descent direction for ELP, i.e. whether (2.2) holds. If all the active functions have the value zero, then the following applies:

THEOREM 3. Suppose that $\hat{f} = 0$, $\hat{V}$ has full rank and Z^TWZ is positive definite. Then p_E, the solution to QP1, is a descent direction for ELP provided it is not zero.

Proof. We have

$$g^Tp_E = \bar{e}^T\bar{p} + \hat{e}^T\hat{p} = \bar{e}^T\bar{\Sigma}\bar{V}^Tp \qquad \text{by (3.1) and (3.2)} .$$

Since $\hat{f} = 0$ we have $p_Y = 0$ and $p = Zp_Z$ as defined by (3.5). Thus

$$g^Tp_E = \bar{e}^T\bar{\Sigma}\bar{V}^TZp_Z = -p_Z^T(Z^TWZ)p_Z \quad \text{by (3.5)} .$$

Since Z^TWZ is positive definite, g^Tp_E must be negative if $p_Z \neq 0$, i.e. $p \neq 0$. □

If $p = 0$, then by (3.5) $Z^T\bar{V}\bar{\Sigma}\bar{e} = 0$ and $x^{(k)}$ is a minimum on the manifold defined by the current active constraints, and hence (4.6) is a consistent set of equations with $\lambda_C = \lambda_L$. Thus either $x^{(k)}$ is the required solution, or one of the components of λ_C is negative or zero, i.e. $|(\pi_C)| \geq 1$ for some j.

If $p = 0$, and at least one of the multipliers is negative, i.e. $|(\pi_C)_j| > 1$, then it is necessary to delete a corresponding function from the active set to obtain a descent direction. If $p = 0$, and the component of π_C with largest modulus is ± 1, corresponding to a zero multiplier, then $x^{(k)}$ may or may not be a solution. We refer to Gill and Murray [10] for the treatment of zero multipliers.

5.1. Nonzero Active Functions

In practice it will rarely be the case that $\hat{f} = 0$, so we now drop this assumption. If we were sufficiently restrictive in the definition of the active set (e.g. no active functions) we could force this condition to be true, but it is important for the efficiency of the algorithm not to be too restrictive in the choice of active set. It could then happen that p_E is an ascent direction for ELP. It is now necessary to introduce some further notation. We denote the components of p_E which correspond to the orthogonal n-vectors Yp_Y and Zp_Z by p_{EY} and p_{EZ} respectively. More specifically, we define:

$$p_{EY} = \begin{bmatrix} Yp_Y \\ \bar{\bar{\Sigma}}\bar{\bar{V}}^T Yp_Y \\ -\hat{\hat{\Sigma}}\hat{\hat{f}} \end{bmatrix} \quad \text{and} \quad p_{EZ} = \begin{bmatrix} Zp_Z \\ \bar{\bar{\Sigma}}\bar{\bar{V}}^T Zp_Z \\ 0 \end{bmatrix} \tag{5.1}$$

We have $p_E = p_{EY} + p_{EZ}$.

Without the assumption that $\hat{f} = 0$, both p_{EY} and p_{EZ} could be ascent directions for ELP. If either component is a descent direction it is possible to take a weignted combination as the search direction. Also, if $Z^T\bar{V}\bar{\Sigma}\bar{e} \neq 0$, the component p_{EZ} can be replaced by the following, which must be a descent direction:

$$\begin{bmatrix} Zq_Z \\ \bar{\Sigma}\bar{V}^T Zq_Z \\ 0 \end{bmatrix}$$

where Zq_Z is the solution to QP3 and q_Z is given by (3.6).

When $Z^T\bar{V}\bar{\Sigma}\bar{e} = 0$ and p_{EY} is an ascent direction for ELP, it is necessary to delete a function from the active set to obtain a descent direction. It is foolish to simply delete the active function with the largest absolute value. The following theorem shows that, in this situation, a constraint can always be found with a negative multiplier estimate. The interpretation of p_{EY} being an ascent direction, regardless of the magnitude of $\| Z^T\bar{V}\bar{\Sigma}\bar{e} \|$, is that too many functions have been selected to be active and are being forced to be approximately zero at $x^{(k)} + p$, thus forcing the inactive functions to increase in modulus more than the active ones are decreasing. Therefore the following is also useful when $Z^T\bar{V}\bar{\Sigma}\bar{e} \neq 0$.

THEOREM 4. Assume $\hat{V}$ has full rank and let p_{EY} be defined by (5.1). If $g^T p_{EY} > 0$, then for some j either $\hat{f}_j > 0$ and $(\pi_C)_j > 1$, or $\hat{f}_j < 0$ and $(\pi_C)_j < -1$.

Proof. It follows from $g^T p_{EY} > 0$ that

$$\bar{e}^T \bar{\Sigma} \bar{V}^T Y p_Y - \hat{e}^T \hat{\Sigma} \hat{f} > 0 .$$

A characterization of the solution of the least squares problem (4.7) is

$$\hat{V} \pi_C = -Y Y^T \bar{V} \bar{\Sigma} \bar{e} .$$

Multiplying both sides by $Y p_Y$, we have from (3.4) that

$$\pi_C^T \hat{f} = \bar{e} \bar{\Sigma} \bar{V}^T Y p_Y$$

and hence

$$\pi_C^T \hat{f} > \hat{e}^T \hat{\Sigma} \hat{f} .$$

Therefore for some j, $(\pi_C)_j \hat{f}_j > |\hat{f}_j|$ and the result follows. □

It also follows from the above that if $g^T p_{EY} = 0$, then either $\hat{f} = 0$ (covered by Theorem 3) or the maximum element of π_C in modulus is one (the minimum multiplier estimate is zero). It is worth noting that Theorem 4 does not hold in general if other multiplier estimates such as λ_L and μ_J are substituted for λ_C.

It follows from Theorem 4 in conjunction with Theorem 2 that if p_{EY} is an ascent direction we can delete the active constraint corresponding to a negative component of λ_C to obtain a first-order feasible direction. As in [15], we will not actually take the steepest descent direction $\tilde{Z} s_{\tilde{Z}}$ in the new null space. Instead, we will first try computing the Newton step $\tilde{Z} q_{\tilde{Z}}$, defined by (4.8) and

$$(\tilde{Z}^T\tilde{W}\tilde{Z})q_{\tilde{Z}} = -\tilde{Z}^T(\bar{V}\bar{\Sigma}\bar{e} + \mathrm{sgn}(\pi_C)_j\hat{v}_j)$$

where the $j^{\underline{th}}$ active function was deleted. If $\tilde{Z}q_{\tilde{Z}}$ is first-order feasible with respect to the deleted constraint, we take as our search direction $\gamma\tilde{Y}p_{\tilde{Y}} + \tilde{Z}q_{\tilde{Z}}$ for some γ, $0 \leq \gamma \leq 1$, where $\tilde{Y}$ spans the range of $\tilde{V}$ and $p_{\tilde{Y}}$ is defined by removing the $j^{\underline{th}}$ equation from (3.4). Otherwise we replace $\tilde{Z}q_{\tilde{Z}}$ by $\tilde{Z}r_{\tilde{Z}}$, where

$$r_{\tilde{Z}} = \begin{bmatrix} q_Z \\ -z^T(\bar{V}\bar{\Sigma}\bar{e} + \mathrm{sgn}(\pi_C)_j\hat{v}_j) \end{bmatrix},$$

q_Z is given by (3.6) and z by (4.8). The proof that $\tilde{Z}r_{\tilde{Z}}$ is a descent direction and is first-order feasible combines the methods of proof of Theorem 5 of [15] and Theorem 2 above in a straightforward way, so we do not present it here.

We do not wish to minimize on manifolds but to delete constraints early when the multiplier estimates become sufficiently reliable. The above discussion of the step to take after deleting a constraint applies to this situation as well as to the situation when $g^Tp_{EY} > 0$. Both λ_C and μ_W are computed when possible so that they can be compared.

5.2. Avoiding a Rank-Deficient Jacobian or an Indefinite Projected Hessian

The methods used to avoid a rank-deficient matrix $\hat{V}$ or an indefinite matrix Z^TWZ are identical to those used in [15]. In the former case, instead of ordering the potential columns of $\hat{A}$ by the size of $\{c_j\}$ we order the potential columns of $\hat{V}$ by the size of $\{|f_j|\}$ in the QR factorization of the Jacobian.

6. Quasi-Newton and Finite Difference Approximations to the Hessian

In practice we may wish to use a quasi-Newton or finite difference approximation to W rather than the analytical Hessians $\{\nabla^2 f_i\}$. The main difference between the ℓ_1 and minimax cases in this regard is that for the ℓ_1 case, W involves Hessians of all the m functions $\{f_i\}$, while in the minimax case W involves the Hessians of only the t active functions. Thus a finite difference approximation to W requires extra gradient evaluations of all m functions at each iteration instead of just those t which are active. Since for many applications (particularly arising from data approximation) m is much larger than n and hence t, this means that finite difference approximations may be considerably more expensive than a quasi-Newton approach for an ℓ_1 problem, while a finite difference approximation may be more efficient for a comparable minimax problem. Note, however, that the extra evaluations of each gradient need only be done $n-t$ times, along each column of Z.

7. Determining the Steplength

We use a special steplength algorithm tailored to the ℓ_1 problem to obtain the steplength α at each iteration. This algorithm is presented in [14]. The initial step α_0 is set to either one, or the shortest estimated step to a zero of an inactive function, if this is less than one. Thus

$$\alpha_0 = \min\{1, \alpha_0'\},$$

where

$$\alpha_0' = \min \left\{ \frac{-\bar{f}_i}{\bar{v}_i^T p} \;\middle|\; \bar{v}_i^T p \neq 0 \quad \text{and} \quad \frac{\bar{f}_i}{\bar{v}_i^T p} < 0 \right\} .$$

8. Selecting the Active Set

Currently we use the same active set strategy as that described for problem $\ell_\infty P$ in [15], with some slight modifications, as follows. We define the scaled function values by $\bar{c}_i = (m|f_i|)/F_1$. Since there may be no active functions the first decision is whether to include the smallest one (in absolute value) in the active set. This decision is made in the same way as the decision of whether to include a second active constraint in the minimax case, replacing the gradient v_1 by the gradient of the ℓ_1 function when no functions are active, i.e. $\bar{V}\bar{\Sigma}\bar{e}$.

9. Relationships to Other Algorithms

We begin this section by discussing algorithms for problem $\ell_1 P$ when the functions $\{f_i\}$ are linear. The equivalent problem ELP is then a linear programming problem, although it is not in standard form. The connection between the linear ℓ_1 problem and linear programming (although using a different formulation from ELP) was observed by Charnes, Cooper and Ferguson [6]. Since then a large number of different methods have been proposed for solving the linear ℓ_1 problem by various linear programming formulations. References to many such methods may be found in [2,3,4,19]. Probably the most widely used method is that of Barrodale and Roberts [3], which solves a variation of ELP put in standard form by a primal simplex method taking account of the special structure. An alternative approach taken by Claerbout and Muir [7] and Bartels, Conn and Sinclair [5] is to solve the problem directly by minimizing the piecewise linear function. However, it is possible to think of these methods as linear programming methods applied to ELP, by considering the connection between the vector w in Bartels, Conn and Sinclair (which corresponds to π in our nonlinear algorithm) and the simplex (Lagrange) multipliers of ELP, just as we do for the nonlinear problem. This is the approach we prefer since it retains the familiar linear programming terminology while avoiding transforming ELP to standard form.

A completely different approach to the linear ℓ_1 problem is to solve a sequence of weighted least squares problems as the Lawson algorithm (see [17]) does for the linear ℓ_∞ problem. This was suggested

by Schlossmacher [18] but Gallant and Gerig [9] show that this method can be unstable.

It is somewhat surprising, given the number of linear ℓ_1 algorithms, that there has been comparatively little work done on the nonlinear ℓ_1 problem. Osborne and Watson [16] solve the nonlinear ℓ_1 problem by solving a sequence of linear ℓ_1 problems. The solution to each linear ℓ_1 problem, obtained by a linear programming technique, is used as a search direction along which the minimum of F_1 is found by an exact line search. We are aware of only two published methods for the nonlinear ℓ_1 problem which use second-order information, both of which appeared only recently. El-Attar, Vidyasagar and Dutta [8] suggest a method related to the penalty method for nonlinear programming in which a sequence of increasingly ill-conditioned unconstrained optimization problems are solved. McLean and Watson [13] propose both a first-order Levenberg-type of method similar to those of [1,12] for the minimax problem, and a method which uses second-order information. The latter is a two-stage method similar to that of Watson [20] for the minimax problem, in which successive linear programming problems are solved until it is thought that the active set has been identified, whereupon a switch is made to solving a system of nonlinear equations by Newton's method. The system has order n+t, since the variables and multipliers are obtained together. Since t may often be close to n (t equals n in the linear case), the systems of equations which are solved may be much larger than the ones we solve.

We are not aware of any published methods of the projected Lagrangian type for problem $\ell_1 P$, although we understand that Bartels and Conn are currently doing some related work. It would be possible to construct a method related to ours but which solves an inequality-constrained quadratic program variant of QP1 at each iteration as Han [11] does for the minimax problem. However, such a QP has n+m variables and it is not possible to transform this directly to an inequality-constrained QP in n variables (as we transform QP1 to QP2). It would be necessary to solve the inequality-constrained QP by a special-purpose method taking into account the special structure, just as the Bartels, Conn and Sinclair method essentially solves the linear program equivalent of ELP by a special-purpose method. See [15] for remarks concerning the relative merits of solving the equality and inequality-constrained QP's.

The remarks on asymptotic local quadratic convergence made in [15] for the minimax problem carry over without difficulty to our algorithm for $\ell_1 P$.

10. Computational Results

We present the results of applying the algorithm to ℓ_1 problems with the same definitions of $\{f_i(x)\}$ as the first four problems presented in [15]. The solutions obtained are listed below.

Problem 1.

$$F_1(\overset{*}{x}) = 0.12434 \quad \text{with} \quad \overset{*}{x} = (0.10094,\ 1.52515,\ 1.97211)^T .$$

Problem 2.

$$F_1(\overset{*}{x}) = 0.0038768 \quad \text{with} \quad \overset{*}{x} = (0.19337,\ 0.19377,\ 0.10893,\ 0.13973)^T$$

Problem 3.

$$F_1(\overset{*}{x}) = 1.00000 \quad \text{with} \quad \overset{*}{x} = (0.0000,\ 0.0002)^T.$$

Problem 4.

$$F_1(\overset{*}{x}) = 7.8942 \quad \text{with} \quad \overset{*}{x} = (0.53597,\ 0.00000,\ 0.031918)^T.$$

The results are summarized in Table 1. The termination conditions were that $\|\hat{f}\|_2 < 10^{-6}$, $\|Z^T\bar{V}\bar{\Sigma}\bar{e}\| < 10^{-6}$, Z^TWZ numerically positive definite and $\lambda_C \geq 0$. The line search accuracy parameter η was set to 0.9 (see [14] for the definition of this parameter). Several other choices of η were tried, but $\eta = 0.9$ was the most efficient, indicating as expected that a slack line search is desirable at least on these problems. The machine used was an IBM 370/168 in double precision, i.e. with 16 decimal digits of accuracy. The column headed NI reports the number of iterations required, which is also the number of times the Hessian was approximated using finite differences. The column headed NF gives the number of function evaluations (not including gradient evaluations for the Hessian approximation).

TABLE 1

	Problem	n	m	$n-t^*$	NI	NF
1	(Bard)	3	15	0	20	20
2	(Kowalik and Osborne)	4	11	0	11	14
3	(Madsen)	2	3	0	15	15
4	(El-Attar et al. #2)	3	6	2	10	11

These results demonstrate that our algorithm can be very efficient. Final quadratic convergence was observed in all cases. The results must however be regarded as preliminary since further work needs to be done regarding the active set strategy.

11. Concluding Remarks

The nonlinear ℓ_1 optimization problem has been shown to be as tractable as the nonlinear minimax problem using a projected Lagrangian algorithm closely related to that of [15]. Although the nonlinearly constrained optimization problem which is equivalent to $\ell_1 P$ involves m extra variables, we have shown how to derive a method which solves successive quadratic programming problems in only n variables. The different roles of multiplier estimates and directions of search in the (n+m)- and n-dimensional spaces have been emphasized.

We could repeat many of the concluding remarks of [15] here. In particular we observe that linear constraints can be incorporated into the algorithm but that nonlinear constraints increase the complexity of problem $\ell_1 P$ to that of the general nonlinear constrained optimization problem. In summary, the method of this paper has been designed to take advantage of all the special properties of the ℓ_1 problem which are not available for general constrained optimization problems.

Acknowledgments:

The authors would like to thank Philip E. Gill, Gene H. Golub, and Margaret H. Wright for a number of helpful discussions.

REFERENCES

[1] Anderson, D.H. and M.R. Osborne (1977). Discrete, nonlinear approximations in polyhedral norms: a Levenberg-like algorithm, Num.Math. 28, pp. 157-170.

[2] Armstrong, R.D. and J.P. Godfrey (1979). Two linear programming algorithms for the linear discrete L_1 norm problem, Math. Comp. 33, pp. 289-300.

[3] Barrodale, I. and F.D.K. Roberts (1973). An improved algorithm for discrete ℓ_1 linear approximation, SIAM J. Num. Anal. 10, pp. 839-848.

[4] Barrodale, I. and F.D.K. Roberts (1978). An efficient algorithm for discrete ℓ_1 linear approximation with linear constraints, SIAM J. Num. Anal. 15, pp. 603-611.

[5] Bartels, R., A.R. Conn and J.W. Sinclair (1978). Minimization techniques for piecewise differentiable functions: The ℓ_1 solution to an overdetermined linear system, SIAM J. Num. Anal. 15, pp. 224-241.

[6] Charnes, A., W.W. Cooper and R. Ferguson (1955). Optimal estimation of executive compensation by linear programming, Manag. Sci. 2, pp. 138-151.

[7] Claerbout, J.F. and Francis Muir (1973). Robust modelling with erratic data, Geophysics 38, pp. 826-844.

[8] El-Attar, R.A., M. Vidyasagar and S.R.K. Dutta (1979). An algorithm for ℓ_1-norm minimization with application to nonlinear ℓ_1-approximation, SIAM J. Num. Anal. 16, pp. 70-86.

[9] Gallant, A.R. and T.M. Gerig (1974). Comments on computing minimum absolute deviations regressions by iterative least squares regressions and by linear programming, Institute of Statistics Rept 911, North Carolina State University.

[10] Gill, P.E. and W. Murray (1977). The computation of Lagrange multiplier estimates for constrained minimization, National Physical Laboratory Rept. NAC 77.

[11] Han, S.P. (1977). Variable metric methods for minimizing a class of nondifferentiable functions, Dept. of Computer Science Rept, Cornell University.

[12] Madsen, K. (1975). An algorithm for the minimax solution of over-determined systems of nonlinear equations, J. Inst. Math. Applics. 16, pp. 321-328.

[13] McLean, R.A. and G.A. Watson (1979). Numerical methods for nonlinear discrete L_1 approximation problems, Proc. Oberwolfach Conf. on Approx. Theory, to appear.

[14] Murray, W. and M. L. Overton (1979). Steplength algorithms for minimizing a class of nondifferentiable functions, Computing 23, pp. 309-331.

[15] Murray, W. and M. L. Overton (1979). A projected Lagrangian algorithm for the nonlinear minimax problem, Systems Optimization Laboratory Report SOL-79-21, Stanford University (submitted for publication in SISSC).

[16] Osborne, M.R. and G.A. Watson (1971). On an algorithm for discrete nonlinear L_1 approximation, Computer J. 14, pp. 184-188.

[17] Rice, J.R. and K.H. Usow (1968). Lawson's algorithm and extensions, Math. Comp. 22, pp. 118-127.

[18] Schlossmacher, E.J. (1973). An iterative technique for absolute derivation curve fitting, J. Amer. Stat. Assoc. 56, pp. 359-362.

[19] Schuette, D.R. (1977). A defense of the Karst algorithm for finding the line of best fit under the L_1 norm, Math. Research Center Rept 1735, University of Wisconsin, Madison.

[20] Watson, G.A. (1979). The minimax solution of an overdetermined system of nonlinear equations, J. Inst. Math. Applics. 23, pp. 167-180.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: SOL 80-4	2. GOVT ACCESSION NO.: AD-A086 162
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): A PROJECTED LAGRANGIAN ALGORITHM FOR NONLINEAR ℓ_1 OPTIMIZATION	5. TYPE OF REPORT & PERIOD COVERED: TECHNICAL REPORT
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Walter Murray and Michael L. Overton	8. CONTRACT OR GRANT NUMBER(s): DAAG29-79-C-0110
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Operations Research - SOL Stanford University Stanford, California 94305	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709	12. REPORT DATE: February, 1980 13. NUMBER OF PAGES: 35
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES: THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

ℓ_1 approximation nonlinear constraints
quadratic programs

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SOL 80-4; by Walter Murray and Michael L. Overton

A PROJECTED LAGRANGIAN ALGORITHM FOR NONLINEAR ℓ_1 OPTIMIZATION

The nonlinear ℓ_1 problem is an unconstrained optimization problem whose objective function is not differentiable everywhere, and hence cannot be solved efficiently using standard techniques for unconstrained optimization. The problem can be transformed into a nonlinearly constrained optimization problem, but it involves many extra variables. We show how to construct a method based on projected Lagrangian methods for constrained optimization which requires successively solving quadratic programs in the same number of variables as that of the original problem. Special Lagrange multiplier estimates are used to form an approximation to the Hessian of the Lagrangian function, which appears in the quadratic program. A special line search algorithm is used to obtain a reduction in the ℓ_1 objective function at each iteration. Under mild conditions, the method is locally quadratically convergent if analytical Hessians are used.

END

DATE
FILMED

8-80

DTIC